DTIC FILE COPY

UNCLASSIFIED

AD-A198 018

(2)

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE:

3. DISTRIBUTION/AVAILABILITY OF REPORT: distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S): Texas A&M Univ., Technical Report #6

5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 88-0845

6a. NAME OF PERFORMING ORGANIZATION: Texas A&M University

6b. OFFICE SYMBOL (If applicable):

7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research

6c. ADDRESS (City, State and ZIP Code): College Station, TX 77853

7b. ADDRESS (City, State and ZIP Code): Same as 8c

8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR

8b. OFFICE SYMBOL (If applicable): NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: ~~AFOSR-F~~F49620-85-C-0144

8c. ADDRESS (City, State and ZIP Code): Bldg 410
Bolling Air Force Base
Washington, DC 20332

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2304	A6	

DTIC ELECTE AUG 2 9 1988 S E

11. TITLE (Include Security Classification): A NOTE ON SECOND ORDER EFFECTS IN A SEMIPARAMETRIC CONTEXT

12. PERSONAL AUTHOR(S): Carroll, Raymond J. and Haerdle, Wolfgang

13a. TYPE OF REPORT: ~~technical~~ Journal

13b. TIME COVERED: FROM 8/87 TO 8/88

14. DATE OF REPORT (Yr., Mo., Day):

15. PAGE COUNT: 16

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): heteroscedastic linear regression model, kernel regression estimators

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

We consider a heteroscedastic linear regression model with normally distributed errors in which the variances depend on an exogenous variable. Suppose that the variance function can be parameterized as $\psi(z_i,\theta)$ with θ unknown. If $\hat{\theta}$ is any root-N consistent estimate of θ based on squared residuals, it is well known that the resulting generalized (weighted) least squares estimate with estimated weights has the same limit distribution as θ were known. The covariance of this estimate can be expanded to terms of order N^{-2}. If the variance function is unknown but smooth, the problem is adaptable, i.e., one can estimate the variance function nonparametrically in such a way that the resulting generalized least squares estimate has the same first order normal limit distribution as if the variance function were completely specified. In a special case we compute an expansion for the covariance in this semiparametric context, and find that the rate of convergence is slower for this estimate than for its parametric counterpart. More importantly, we find that there is an effect due to how well one estimates the variance function. We use a kernel regression

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL: Major Brian Woodruff

22b. TELEPHONE NUMBER (Include Area Code): (202) 767-5026

22c. OFFICE SYMBOL: NM

DD FORM 1473, 83 APR   EDITION OF 1 JAN 73 IS OBSOLETE.   88 8 25 093

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

AFOSR-TR- 88-0845

A NOTE ON SECOND ORDER EFFECTS
IN A SEMIPARAMETRIC CONTEXT

Technical Report # 6

Raymond J. Carroll and Wolfgang Haerdle

A NOTE ON SECOND ORDER EFFECTS IN A SEMIPARAMETRIC CONTEXT

Raymond J. Carroll

Wolfgang Haerdle

Raymond J. Carroll is with the Department of Statistics, Texas A&M University. His research was supported by the Air Force Office of Scientific Research contract [redacted]-F-49620-85-C-0144. Wolfgang Haerdle is with the Institut fuer Wirtschaftstheorie II, Adenauerallee 24-26, Universitaet Bonn, D-5300 Bonn 1, West Germany. His research was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303.

ABSTRACT

We consider a heteroscedastic linear regression model with normally distributed errors in which the variances depend on an exogenous variable. Suppose that the variance function can be parameterized as $\psi(z_i, \theta)$ with θ unknown. If $\hat{\theta}$ is any root-N consistent estimate of θ based on squared residuals, it is well known that the resulting generalized (weighted) least squares estimate with estimated weights has the same limit distribution as if θ were known. The covariance of this estimate can be expanded to terms of order N^{-2}. If the variance function is unknown but smooth, the problem is adaptable, i.e., one can estimate the variance function nonparametrically in such a way that the resulting generalized least squares estimate has the same first order normal limit distribution as if the variance function were completely specified. In a special case ~~we~~ compute an expansion for the covariance in this semiparametric context, and ~~find that~~ the rate of convergence is found to be slower for this estimate than for its parametric counterpart. More importantly, ~~we find that~~ there is an effect due to how well one estimates the variance function. ~~We use~~ a kernel regression estimator~~, and find that~~ the optimal bandwidth in ~~our~~ problem is of the usual order, but ~~that~~ the constant depends on the variance function as well as the particular linear combination being estimated.

SECTION 1 : Introduction

We consider a heteroscedastic linear regression model with normally distributed errors and replication :

$$y_{ij} = x_i^t\beta + \sigma_i\eta_{ij} \ , \ (i = 1, \ \ldots \ ,N \ ; \ j = 1,2) \ ;$$

$$\sigma_i^2 = \psi(z_i,\theta) \ ; \tag{1.1}$$

$$E(\eta_{ij}) = 0 \quad ; \quad \text{Variance}(\eta_{ij}) = 2.$$

In this model, the regression parameter is β, and the variance function is ψ. The $\{z_i\}$ are scalars, possibly a component of the p-dimensional vectors $\{x_i\}$. Throughout, we will assume that the $\{x_i,z_i\}$ are independent and identically distributed random variables mutually independent of the $\{\eta_{ij}\}$. The errors $\{\eta_{ij}\}$ are assumed to be independent normally distributed random variables. The reason that the variance of η_{ij} equals 2 will become clear later.

Let $\hat{\theta}$ be the mle of θ. The mle $\hat{\beta}_w$ of β is a generalized least squares estimate, i.e., weighted least squares with the estimated weights $1/\psi(z_i,\hat{\theta})$. Let

$$S_N = N^{-1}\sum_{i=1}^{N} x_i \ x_i^t \ / \ \psi(z_i,\theta) \longrightarrow S \text{ (positive definite)}, \tag{1.2}$$

then it is well known that $\hat{\beta}_w$ is asymptotically normally distributed with mean β and covariance S^{-1}/N, i.e., with "$\Rightarrow$" denoting convergence in distribution,

$$N^{1/2}(\hat{\beta}_w - \beta) \Rightarrow \text{Normal}(0,S^{-1}). \tag{1.3}$$

The limit distribution (1.3) is the same as if θ were known, so that (1.3) expresses a parametric adaptation result.

A simplification of an argument of Rothenberg (1984) shows that $\hat{\beta}_w$ is

symmetrically distributed about β with a covariance expansion

$$(1.4) \qquad \text{Covariance}\left[N^{1/2}(\hat{\beta}_w - \beta) \right] = S^{-1} + N^{-1} \Lambda_w + o(N^{-1}),$$

where Λ_w is a positive definite matrix. Such second order covariance expansions when the variances depend on the mean and/or the errors are not normally distributed have been investigated by Carroll, Wu & Ruppert (1987).

Suppose that instead of a parametric model, the form of the variance function is not known a priori, so that we can write

$$(1.5) \qquad \sigma_i^2 = \psi(z_i) = 1 / g(z_i), \ \psi \text{ unknown}.$$

Now the unknown parameters are (β,ψ), so we are in a semiparametric context, see Bickel (1982) and Begun, et al. (1983). It is easy to show that the semiparametric information bound here is the same as if ψ were known. Carroll (1982), Robinson (1986) and Carroll, Ruppert & Stefanski (1987) have constructed adaptive estimates as follows. By smoothing techniques such as kernel or nearest neighbor regression, they form an estimate $\hat{\psi}$ of ψ, and then construct the generalized least squares estimate $\hat{\beta}_g$ of β with the estimated weights $1/\hat{\psi}(z_i)$. These estimates have the same limit distribution as if ψ were known, i.e.,

$$N^{1/2}(\hat{\beta}_g - \beta) \Rightarrow \text{Normal}(0,S^{-1}) \ .$$

If $\hat{\psi}$ is chosen appropriately, $\hat{\beta}_g$ is symmetrically distributed about β. In this paper, we pick a particular estimate $\hat{\psi}$ based on kernel regression techniques and compute an analogue to the covariance expansion (1.4), namely

$$\text{Covariance}\left[N^{1/2}(\hat{\beta}_g - \beta) \right] = S^{-1} + N^{-4/5}\Lambda_g + o(N^{-1}).$$

There are two major conclusions. The first is that the second order covariance expansion converges at a slower rate for the semiparametric model (1.L) than it does for the parametric model (1.1). Of more general interest is that the optimal bandwidth for estimating any linear combination of the regression parameter β is still of the usual order, but it depends not only on the variance function but also on the particular linear combination being estimated.

In a sense the context we are working in is narrow, but there are some general implications to our results. In the semiparametric context, there is some concern that much larger sample sizes than usual will be needed to achieve approximate normality than is true in a parametric model. Hsieh & Manski (1987) state "It is sometimes asserted that satisfactory nonparametric estimation of score functions requires very large samples; hence, adaptive estimates should perform poorly in moderate size samples".

Our results indicate that semiparametric adaptive estimates should indeed converge more slowly than do parametric estimates, which is not too surprising a result but is at least worth nailing down. With considerable fine tuning of their estimate of the nonparametric part of their model, Hsieh & Manski are able to do fairly well in their two-sample problem. It is clear from their simulations that how well one estimates the nonparametric part of their semiparametric model can affect the small sample properties of the parametric estimator. Our results are a theoretical complement to their simulations. How well one estimates the semiparametric nuisance function ψ can affect the small sample performance of the parametric estimates.

SECTION 2 : A Second Order Covariance Expansion

The key to our construction is that replication in a normally distributed context allows us to do weighted least squares with estimated weights which are distributed independently of the "data". Let

$$\epsilon_i = (\eta_{i1} + \eta_{i2})/2 \quad ; \quad \delta_i = \epsilon_i \, \psi^{1/2}(z_i) \ ;$$

$$\epsilon_{i*} = (y_{i1} - y_{i2})/2 = \psi^{1/2}(z_i) \ (\eta_{i1} - \eta_{i2})/2 \ .$$

Note that the sequence $\{\epsilon_{i*}\}$ is observable. Because the $\{\eta_{ij}\}$ are normally distributed, the sequences $\{\delta_i\}$ and $\{\epsilon_{i*}\}$ are mutually independent and identically distributed standard normal random variables. Also, the $\{\epsilon_i\}$ are distributed independently of the $\{\epsilon_{i*}\}$. Since

$$E\{ \epsilon_{i*}^{\,2} \} = \psi(z_i) \ ,$$

it is plausible to base estimates of the weights on the $\{\epsilon_{i*}^{\,2}\}$. Of course, this will not be the most efficient way to estimate the variance function ψ, but will still allow us to estimate β efficiently to first order. We first write the results in terms of $g(z) = 1/\psi(z)$, see (1.5). Let $\hat{g}_N$ be an estimate of g which is based solely on the $\{\epsilon_{i*}^{\,2}\}$. Make the following definitions :

$$\hat{v}_N(z) = \hat{g}_N(z) - g(z) \quad ;$$

$$S_N = N^{-1} \sum_{i=1}^{N} x_i x_i^{\,t} \, g(z_i) \quad ; \quad \hat{S}_N = N^{-1} \sum_{i=1}^{N} x_i x_i^{\,t} \, \hat{g}_N(z_i) \quad ;$$

$$R_N = N^{-1/2} \sum_{i=1}^{N} x_i \delta_i g(z_i) \quad ; \quad \hat{R}_N = N^{-1/2} \sum_{i=1}^{N} x_i \delta_i \, \hat{g}_N(z_i) \ ;$$

$$M_N = \hat{S}_N - S_N \quad ; \quad Q_N = \hat{R}_N - R_N \ .$$

LEMMA 1 : For any N,

$$\mathrm{Cov}\left[N^{1/2}(\hat{\beta}_g - \beta) \right] = S_N^{-1} + E\left[T_N \, T_N{}^t \right]$$

PROOF OF LEMMA 1 : Note that

$$T_N = N^{1/2}(\hat{\beta}_g - \hat{\beta}_w) \; : \; E\{T_N\} = 0 \; .$$

the latter following easily since $\hat{g}_N$ is independent of the $\{\delta_i\}$. Since the distribution of T_N does not depend on β and $\hat{\beta}_w$ is a complete sufficient statistic for β, by Basu's Lemma T_N is independent of $\hat{\beta}_w$. This means that

$$\mathrm{Cov}\left[N^{1/2}(\hat{\beta}_g - \beta) \right] = \mathrm{Cov}\left[N^{1/2}(\hat{\beta}_w - \beta) \right] + \mathrm{Cov}\left[T_N \right]$$

$$= S_N{}^{-1} + E\left[T_N \, T_N{}^t \right] . \qquad \square$$

Because $T_N \, T_N{}^t$ is positive (semi) definite, Lemma 1 implies that estimating weights by our method results in an inflation in variance. Define

$$\mathbb{C}_N = T_N \, T_N{}^t - (Q_N - M_N R_N) \, (Q_N - M_N R_N)^t$$

We show in the appendix that under reasonable regularity conditions, $\mathbb{C}_N = o_p(N^{-1})$. Thus, it is not too implausible to assume that

(2.1) $$N \, E\left[\mathbb{C}_N \right] \longrightarrow 0 \quad \text{as } N \longrightarrow \infty,$$

THEOREM 1 : Assume (2.1). Then

(2.2) $$\mathrm{Cov} \left[N^{1/2}(\hat{\beta}_g - \beta) \right]$$

$$= S_N^{-1} + S_N^{-1}\Big[\ E\{\ Q_N\ Q_N^{\ t}\ \} - E\{\ M_N\ S_N^{-1}\ M_N^{\ t}\ \}\ \Big]S_N^{-1} + o(N^{-1}).$$

The proof is in the appendix. □

The translation from $\hat{g}_N(z) - g(z)$ in the definition of Q_N and M_N is to note that

$$(2.3) \qquad \hat{g}_N(z) - g(z) \cong -\Big[\ \hat{\psi}_N(z) - \psi(z)\ \Big] \,/\, \psi^2(z)\ .$$

We will ignore the error in (2.3) by subsuming it under "additional regularity conditions". Define

$$v_i = x_i\ x_i^{\ t}\ \psi(z_i)\ ;$$

$$A_N = N^{-1}\sum_{i=1}^{N} v_i\ \Big[\ \hat{\psi}_N(z) - \psi(z)\ \Big]^2 \,/\, \psi^4(z)\ ;$$

$$B_N = N^{-1}\sum_{i=1}^{N} x_i\ x_i^{\ t}\ \Big[\ \hat{\psi}_N(z) - \psi(z)\ \Big] \,/\, \psi^2(z)\ .$$

The direct translation from (2.2) is

$$(2.4) \qquad \mathrm{Cov}\ \Big[\ N^{1/2}(\hat{\beta}_g - \beta)\ \Big]$$

$$= S_N^{-1} + S_N^{-1}\Big[\ E\{\ A_N\ \} - E\{\ B_N\ S_N^{-1}\ B_N^{\ t}\ \}\ \Big]S_N^{-1} + o(N^{-1}).$$

In the next section we compute (2.4) in a special case.

SECTION 3 : An Example

The purpose of this section is to get an explicit expression for the covariance expansion (2.4) in a special case. The major question is whether the second term on the right hand side is of order $O(N^{-1})$ as it would be in the parametric case. We use a kernel regression estimator. Let $K(\cdot)$ be a symmetric density with bounded support and let $f(\cdot)$ be the marginal density of the $\{z_i\}$. Let $b = b_N \longrightarrow 0$ be the bandwidth and define

$$K_b(u) = b^{-1} K(u/b) \ ;$$
$$[K] = \int u^2 K(u) \, du \ .$$

The estimator we will use is a leave-one-out type estimator, which Robinson (1986) has also found to be convenient analytically:

$$\hat{\psi}_N(z_i) = \sum_{j \neq i}^{N} \epsilon_{j*}^2 K_b(z_j - z_i) \Big/ \sum_{j \neq i}^{N} K_b(z_j - z_i).$$

We will also need the following :

(3.1) $$N b_N^6 \longrightarrow 0 \ .$$

(3.2) $$\int v K^2(v) \, dv = 0 \ .$$

(3.3) If ψ_j is the jth derivative of ψ and f_1 is the first derivative of f, then

$$c_1(v) = (1/2) \, d(v) \, [K], \text{ where}$$
$$d(v) = \psi_2(v) + 2 f_1(v) \psi_1(v) / f(v).$$

Define the following terms :

$$c_K^{(1)} = \int K^2(v) \, dv \ ; \qquad \mu_4(v) = E\left[\{\epsilon_{j*}^2 - \psi(v)\}^2 \mid z{=}v \right]$$

$$c_2(v) = c_K^{(1)} \mu_4(v) / f(v)$$

$$d_1 = S_N^{-1} E\left[xx^t c_2(z)\psi^{-3}(z)\right] S_N^{-1} \; ; d_2 = S_N^{-1} E\left[xx^t c_1^{\,2}(z)\psi^{-3}(z)\right] S_N^{-1}$$

$$\tau = S_N^{-1} E\left[xx^t c_1(z)\psi^{-2}(z)\right] S_N^{-1} \; ; \quad S = E\left[xx^t/\psi(z)\right] \quad \text{(see (1.2))}$$

$$d_3 = d_2 - \tau\,\tau^t \geq 0 \text{ (by Cauchy Schwarz).}$$

<u>THEOREM 2</u> : Assume (2.4). Then, for estimating any linear combination $a^t\beta$,

$$\text{(3.4)} \qquad \text{Cov}\left[N^{1/2}(a^t\,\hat{\beta}_g - a^t\beta) \right]$$

$$= a^t \left[S^{-1} + (Nb_N)^{-1} d_1 + b_N^{\,4}\, d_3\right] a + o(b_N^4 + (Nb_N)^{-1}).$$

From (3.4), the optimal bandwidth is $cN^{-1/5}$, where

$$\text{(3.5)} \qquad c = \left[q\; a^t d_1 a\right]^{1/5} / \left[\,4\; a^t\; d_3\; a\,\right]^{1/5} . \qquad \square$$

Note how the optimal bandwidth depends on the design and which linear combination you are interested in estimating.

<u>Remark</u> : Assume the result of Lemma 1, we could generalize (3.4) and (3.5) somewhat by allowing the z_i to be q-vectors. The changes needed are these:

$$K_b(u) = b^{-q}K(u/b); \quad [K] = \int uu^t K(u)du \; ;$$

$$c_1(v) = (1/2)\ \text{trace}(d(v)\ [K]), \text{ where}$$

$$d(v) = \psi_2(v) + (\psi_1(v)f_1(v)^t + f_1(v)\psi_1(v)^t)/f(v).$$

The optimal bandwidth is $cN^{-1/(4+q)}$ and (3.4) and (3.5) become

$$\text{(3.4)}^* \qquad \text{Cov}\left[N^{1/2}(a^t\hat{\beta}_g - a^t\beta) \right]$$

$$= a^t \left[S^{-1} + (Nb_N{}^q)^{-1} d_1 + b_N{}^4\, d_3\right] a + o(b_N^4 + (Nb_N{}^q)^{-1}).$$

(3.5)* $$c = \left[q\ a^t d_1 a\ \right]^{1/(4+q)} \Big/ \left[\ 4\ a^t\ d_3\ a\ \right]^{1/(4+q)} .$$

REFERENCES

Begun, J. M., Hall, W. J., Huang, W. M. & Wellner, J. A. (1983). Information and asymptotic efficiency in parametric - nonparametric models. Annals of Statistics 11, 432-452.

Bickel, P. J. (1982). On adaptive estimation. Annals of Statistics 10, 647-671.

Carroll, R. J. (1982). Adapting for heteroscedasticity in linear models, Annals of Statistics 10, 1224-1233.

Carroll, R. J. & Ruppert, D. (1987). Transformations And Weighting In Regression. Chapman & Hall, London & New York.

Carroll, R. J., Ruppert, D., & Stefanski, L. A. (1987). Adaptating for heteroscedasticity in regression models. University of North Carolina Technical Report #1702.

Carroll, R. J., Wu, C. F. J. & Ruppert, D. (1987). Variance expansions and the bootstrap in generalized least squares. Preprint.

Collomb, G. (1977). Quelques proprietes de la methode du noyau pout l'estimation non parametrique de la regression en un point fixe. Comptes Rendu Acad. Sc. Paris 285, 289-292.

Collomb, G. (1981). Estimation non-parametrique de la regression: revue bibliographique. International Statistical Review 49, 75-93.

Robinson, P. M. (1986). Asymptotically efficient estimation in the presence of heteroscedasticity of unknown form. Econometrica, to appear.

Rothenberg, T. J. (1984). Approximate normality of generalized least squares estimates. Econometrica 52, 811-825.

APPENDIX A

Without loss of generality, we may set $S_N = I$.

LEMMA A.1 : Assume that

(A.1) $$\| M_N \|^3 = o_p(N^{-1}) .$$

(A.2) $$\| Q_N \| \xrightarrow{p} 0,$$

(A.3) $$\| M_N M_N Q_N \| = o_p(N^{-1}) .$$

Then,

$$T_N T_N{}^t = (Q_N - M_N R_N)(Q_N - M_N R_N)^t + o_p(N^{-1}) . \qquad \square$$

PROOF OF LEMMA A.1 : We have that

$$\hat{S}_N{}^{-1} - S_N{}^{-1} = -M_N + M_N M_N + O_p(\| M_N \|^3) .$$

Thus,

$$T_N = \hat{S}_N{}^{-1} \hat{R}_N - S_N{}^{-1} R_N = (\hat{S}_N{}^{-1} - S_N{}^{-1})\hat{R}_N + S_N{}^{-1}(\hat{R}_N - R_N)$$

$$= (\hat{S}_N{}^{-1} - S_N{}^{-1})R_N + (\hat{S}_N{}^{-1} - S_N{}^{-1})Q_N + S_N{}^{-1} Q_N$$

$$= \left[-M_N + M_N M_N \right] R_N + \left[-M_N + M_N M_N \right] Q_N + S_N{}^{-1} Q_N + o_p(N^{-1})$$

$$= S_N{}^{-1} Q_N - M_N R_N + M_N M_N R_N - M_N Q_N + o_p(N^{-1}) ,$$

the last following from (A.1) - (A.3). From (A.1) - (A.3), we have

$$T_N T_N{}^t = S_N{}^{-1} Q_N Q_N{}^t S_N{}^{-1} - M_N R_N Q_N{}^t S_N{}^{-1}$$

$$- \left[M_N R_N Q_N{}^t S_N{}^{-1} \right]^t + M_N R_N R_N{}^t M_N + o_p(N^{-1}).$$

Since $S_N = I$, the proof is complete. $\square$

LEMMA A.2 : Define

$$v_i = x_i x_i^t / g(z_i) = x_i x_i^t \psi(z_i).$$

Then,

$$E\left[Q_N Q_N^t \right] = N^{-1} \sum_{i=1}^{N} E\left[v_i \{\hat{g}_N(z_i) - g(z_i)\}^2 \right] . \qquad \square$$

PROOF OF LEMMA A.2 : Since

$$Q_N Q_N^t = N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \delta_i \delta_i x_i x_j^t \{\hat{g}_N(z_i) - g(z_i)\} \{\hat{g}_N(z_j) - g(z_j)\} ,$$

and since $\hat{g}_N$ and the $\{\delta_i\}$ are independent, we find that

$$E\left[Q_N Q_N^t \,\middle|\, \hat{g}_N, \{x_i, z_i\} \right] = N^{-1} \sum_{i=1}^{N} v_i \{\hat{g}_N(z_i) - g(z_i)\}^2 .$$

This completes the proof. $\square$

LEMMA A.3 : We have that

$$E\left[(Q_N - M_N R_N) (Q_N - M_N R_N)^t \right] = E\left[Q_N Q_N^t \right] - E\left[M_N M_N^t \right] .$$

PROOF OF LEMMA A.3 : Exploiting the independence of $\hat{g}_N$ and the $\{\delta_i\}$, as well as remembering that $S_N = I$, we see that

$$E\left[M_N R_N R_N^t M_N^t \right] = E\left[M_N M_N^t \right] .$$

It thus suffices to show that

$$E\left[M_N R_N Q_N^t \mid \hat{g}_N, \{x_i, z_i\} \right] = M_N M_N^t .$$

This is routine. □

PROOF OF THEOREM 1 : The proof follows from the previous Lemmas. □

APPENDIX B

Our calculations rely on the following result due to Collomb (1977, 1981).

PROPOSITION B : Let K be a symmetric q-dimensional density. Let $(X,Y) \in \mathbb{R}^{q+1}$, where X is a q-vector. Define $m(x) = E[y \mid X=x]$ and let $f(x)$ be the marginal density of X. Assume that $m(x)$ and $f(x)$ are four times continuously differentiable. Define

$$K_b(u) = b^{-q}K(u/b)$$

$$\hat{m}_b(u) = \sum_{i=1}^{N} Y_i K_b(x-X_i) \Big/ \sum_{i=1}^{N} K_b(x-X_i)$$

$$v(x) = E[\{Y - m(x)\}^2 \mid X=x]$$

$$[K] = \int u\, u^t\, K(u)\, du$$

$$b(x) = (1/2)m_2(x) + (1/2)[m_1(x)f_1{}^t(x) + f_1(x)m_1{}^t(x)]/f(x).$$

Then, as $b \longrightarrow 0$,

$$E\hat{m}_b(x) - m(x) = b^2 \operatorname{trace}(b(x)\,[K]) + O(b^4)\ ;$$

$$\operatorname{Var}\{\hat{m}_b(x)\} = (Nb^q)^{-1}\, v(x) \int K^2(u)\, du \,/\, f(x) + O(N^{-1}b^{-q+2}).$$ □

LEMMA B.1 : Under regularity conditions (Proposition B) on ψ and K,

$$s(z_i) = E\left[\hat{\psi}_N(z_i) \;\middle|\; z_i\right] \tag{B.1}$$

$$= \psi(z_i) + b_N{}^2\, c_1(z_i) + O_p(b_N{}^4)\ .$$

PROOF OF LEMMA B.1 : Immediate from Proposition B. □

LEMMA B.2 : If $s(z_i)$ is defined in (B.1), we have

$$E\left[[\hat{\psi}_N(z_i) - s(z_i)]^2 \;\middle|\; z_i \right]$$

$$= c_2(z_i)/\{Nb_n^{\ q}\} + O_p(N^{-1}b_N^{\ 2-q}) \ .$$

PROOF OF LEMMA B.2 : From Proposition B with $Y_i = \epsilon_{i*}^2$, $X_i = z_i$, $m(x) = \psi(x)$. □

LEMMA B.3 : We have

$$E\left[[\hat{\psi}_N(z_i) - \psi(z_i)]^2 \;\middle|\; z_i \right]$$

$$= c_2(z_i)/\{Nb_N^{\ q}\} + b_n^{\ 4}\ c_1^{\ 2}(z_i) + O_p(b_N^{\ 6} + N^{-1}b_N^{\ (2-q)}).$$

PROOF OF LEMMA B.3 : The expansion in question is

$$E\left[[\hat{\psi}_N(z_i) - s(z_i)]^2 \;\middle|\; z_i \right] + [s(z_i) - \psi(z_i)]^2$$

$$= c_2(z_i)/\{Nb_N\} + b_n^{\ 4}\ c_1^{\ 2}(z_i) + O_p(N^{-1}b_N^{\ 2-q}) + O_p(b_N^{\ 6}) \ . \qquad \square$$

Now assume without loss that $S_N = I$. Define

$$W_i = x_i\ x_i^{\ t} \ / \ \psi^3(z_i) \ ; \ D_i = \psi(z_i)\ W_i \ .$$

LEMMA B.4 : As $N \longrightarrow \infty$,

$$E\left[N^{-1} \sum_{i=1}^{N} W_i\ [\hat{\psi}_N(z_i) - \psi(z_i)]^2 \right] \tag{B.2}$$

$$= d_1/\{Nb_N^{\ q}\} + b_N^{\ 4}\ d_2 + O(b_N^{\ 6} + N^{-1}b_N^{\ (2-q)}) \ .$$

PROOF OF LEMMA B.4 : Apply Lemma B.3 after conditioning on (x_i, z_i). □

Lemma B.4 gives us the form of $E\{A_N\}$ in (2.4), and in order to complete the calculation we note that $B_N = B_{N1} + B_{N2}$, where

$$B_{N1} = N^{-1} \sum_{i=1}^{N} D_i \, [\hat{\psi}_N(z_i) - s(z_i)]$$

$$B_{N2} = N^{-1} \sum_{i=1}^{N} D_i \, [\psi(z_i) - s(z_i)] \; .$$

It is easy to see by conditioning that

$$E\{B_{N1} \, B_{N2}\} = 0 \; .$$

LEMMA B.5 : As $N \longrightarrow \infty$,

$$E\{ B_{N2} \, B_{N2}{}^t\} = b_N{}^4 \, \tau \, \tau^t + o(N^{-1} b_N{}^q + b_n{}^4) \; .$$

PROOF OF LEMMA B.5 : Let $c_* = E\{ B_{N2} \}$. Now,

$$c_* = b_N{}^2 \, \tau + O(b_N{}^4)$$

By (3.1) and (B.1),

$$E\{ B_{N2} \, B_{N2}{}^t\} = E\Big[[B_{N2} - c_*] \, [B_{N2} - c_*]^t \Big] + c_* \, c_*{}^t$$

$$= c_* \, c_*{}^t + o(N^{-1}) = b_N{}^4 \, \tau \, \tau^t + o(N^{-1} b_N{}^q + b_n{}^4). \quad □$$

It is the calculation of the second moment matrix for B_{N1} that causes the most difficulties. Write $D(z_i) = D_i$.

<u>LEMMA B.6</u> : As $N \longrightarrow \infty$,

$$E\{ B_{N1} B_{N1}{}^t\} = N^{-1} E\Big[\mu_4(z) D(z) D(z)^t \Big]$$
$$- N^{-1} E\Big[D(z) \psi(z) \Big] E\Big[D(z) \psi(z) \Big]^t + o(N^{-1}) = O(N^{-1}) .$$

<u>PROOF OF LEMMA B.6</u> : B_{N1} is the average of mean zero but not independent random variables. By Lemma B.2, we have

$$E\{ B_{N1} B_{N1}{}^t\} = N^{-2} \sum_{i=1}^{N} \sum_{j\neq i}^{N} E\Big[D_i D_j \xi_i \xi_j \Big] + o(N^{-1}),$$

$$\text{where } \xi_i = \hat{\psi}_N(z_i) - s(z_i) .$$

By a direct calculation,

(B.3)
$$E\{ \xi_i \xi_j \mid z_i, z_j \} =$$

$$= (N-1)^{-2} \sum_{k\neq i}^{N} \sum_{m\neq j}^{N} E\Big[[\kappa_N(i,k) - s(z_i)] [\kappa_N(j,m) - s(z_j)] \mid z_i, z_k\Big].$$

$$\text{where } \kappa_N(i,k) = \epsilon_{k*}{}^2 K\{(z_k - z_i)/b_N\} / \{b_N f_Z(z_i)\} ;$$
$$s(z_i) = E\Big[\kappa_N(i,k) \Bigm| z_i \Big] .$$

If (i,j,k,m) are all distinct, the expectation is zero. There is only one case that $(k=j) \neq (i=m)$, and its contribution is of order N^{-2}, which is too small to matter. There are $(N-2)$ terms in which $k=m\neq i$, $k=m\neq j$, $i\neq j$. Thus, (B.3) is

$$N^{-1}\int \Big[\mu_4(v) K\{(v-z_i)/b_N\} K\{(v-z_j)/b_N\} f_Z(v) \Big]$$
$$\times \Big[b_N{}^2 f_Z(z_i) f_Z(z_j) \Big]^{-1} dv - N^{-1} \psi(z_i)\psi(z_j) + o_p(N^{-1})$$

$$= N^{-1} \left[H_N(z_i, z_j) - \psi(z_i)\psi(z_j) \right] + o_p(N^{-1}).$$

We thus see that

$$E\{ B_{N1} B_{N1}^{t} \} = N^{-1} E\left[D(z_i)\, D(z_j)\, \{ H_N(z_i, z_j) - \psi(z_i)\, \psi(z_j) \right] + o(N^{-1}) .$$

To complete the proof we have to show that

$$\text{(B.4)} \qquad E\left[D(z_i)\, D(z_j)\, H_N(z_i, z_j) \right] = O(1) .$$

Taking into account the form of H_N, we find that (B.4) equals

$$\int\int\int b_N^{-2}\, D(z_i)\, D(z_j)\, \mu_4(v)\, K\{(v - z_i)\}\, K\{(v - z_j)\} f_Z(v)\, dv\, dz_i\, dz_j$$

$$= \int\int\int \mu_4(v)\, K(w_i)\, K(w_j)\, D(v + b_N w_i)\, D(v + b_N w_j)\, f_Z(v)\, dw_i\, dw_j\, dv$$

$$= \int \mu_4(v)\, D(v)\, D(v)\, f_Z(v)\, dv + O(b_N^{2}) ,$$

completing the proof. □